Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 134 (2017) 865–874

Procedia Engineering

www.elsevier.com/locate/procedia

9th International Conference on Sustainability in Energy and Buildings, SEB-17, 5-7 July 2017, Chania, Crete, Greece

Mining typical load profiles in buildings to support energy management in the smart city context

Alfonso Capozzoli[a]*, Marco Savino Piscitelli[a], Silvio Brandi[a]

[a]*Politecnico di Torino, DENERG, TEBE Research Group, corso Duca degli Abruzzi 24, Turin, 10129, Italy*

Abstract

Mining typical load profiles in buildings to drive energy management strategies is a fundamental task to be addressed in a smart city environment. In this work, a general framework on load profiles characterisation in buildings based on the recent scientific literature is proposed. The process relies on the combination of different pattern recognition and classification algorithms in order to provide a robust insight of the energy usage patterns at different levels and at different scales (from single building to stock of buildings). Several implications related to energy profiling in buildings, including tariff design, demand side management and advanced energy diagnosis are discussed. Moreover, a robust methodology to mine typical energy patterns to support advanced energy diagnosis in buildings is introduced by analysing the monitored energy consumption of a cooling/heating mechanical room.

Keywords: buiding energy management; building energy profiling; load profiles characterisation; typical energy patterns.

1. Introduction

The last decades have seen a profound transformation in the energy system of many countries worldwide. The progressive introduction of renewable energy sources in buildings in a smart city context and the consequent effort for decarbonisation have changed the way to use and manage energy. The improvement of building energy efficiency is a central theme both in scientific community and global political environment. Important opportunities to address

* Alfonso Capozzoli. Tel.: +39-011-090-4413.
E-mail address: alfonso.capozzoli@polito.it

10.1016/j.egypro.2017.09.545

this task are provided by the implementation of strategies aimed at enhancing the building energy management and operation. In this context the increasing implementation of Information and Communication Technologies (ICT) and Energy Management System (EMS) in buildings has enabled an easier availability of a huge amount of heterogeneous building related data making it possible a bi-directional communication between infrastructures and operators [1]. As a consequence, in the last few years, an increasing interest of researchers and energy companies in exploring the potentialities of data analytics procedures to extract information on the actual building energy behavior was observed [2–6]. The application of data analytics techniques coupled with a robust physics-based expertise can effectively support the implementation of procedures or strategies aimed at enhancing the operational performance of buildings [7]. In particular, the mining of time-series data has recently gained high attention in the scientific literature as a way to describe the load patterns and the boundary conditions (e.g., weather, time period or user/customer features) influencing their particular variation over time. The electrical or thermal load time series are usually characterised by a particular trend with stochastic components and time based cycle at annual, seasonal and daily scales [8].

In the process of load profiles characterisation, data driven recognition techniques play a key role for the identification of typical operational patterns and trends in a high-dimensional time series [9]. Pattern recognition procedures usually require time series to be segmented in a pre-processing step. In the analytical process of building load profiles, time series are usually chunked into subsequences through a fixed length window to obtain a time scale based profiles. In the majority of energy and buildings applications load profiles are usually well described on a daily scale. The mining of time based load profiles is an emergent task which enables the implementation of various energy management and diagnostic strategies at both single and multiple buildings/customers level. Load profiles can be analysed at different scales and/or at different levels of detail or for specific building energy services. The process of daily load profiling primarily consists in grouping similar load profiles using domain expert based procedures, statistical methods or data mining algorithms. For each group of similar load profiles a representative load pattern can be extracted. The shape of a load profile is usually representative of an operational pattern of a building/customer or a sub-system. It is intended as the daily-based set of time-related features defining that specific pattern under bounded load conditions. When time series sub-sequences are compared to discover similar shapes a data scaling is a task to carry out in order to isolate the effect of the magnitude and allow the data mining algorithms to perform correctly for this purpose. Clustering algorithms proved to be particularly effective in discovering robust energy patterns from time series data [9]. In the framework of energy profiling data mining techniques are commonly applied to classify energy consumers or to detect anomalous energy patterns or for the short-term forecasting according to customer/building and temporal variables. In fact, the mining of typical load patterns can be also viewed as a powerful pre-processing phase for the implementation of predictive models of building energy consumption [8]. For different typologies of users (e.g., industrial, residential, commercial) relevant differences could be observed in the daily energy profiles. Furthermore, for the same typology of users, various daily load profiles can be extracted usually depending by the weather, the day of the week, the season or specific features of the building/customer. Depending on whether a single or a group of buildings/customers are analysed, different implications arise from the process of load profiles characterisation. In the first case a detailed diagnostic analysis of energy time series is performed to discover interesting energy patterns characterising the operation of building/customer or sub-systems. In the second case, instead, the objective becomes a load classification to discover typical classes of buildings/customers according to profile shape similarity [10,11]. The process of typical load profiles recognition is beneficial for different actors in the smart city environment:

- An in-depth knowledge of building typical load profiles could help managers in developing different strategies involving energy savings opportunities related to the management of renewable energy systems [12] or thermal/electrical storage systems. Information on typical daily patterns may drive the design of targeted tariff plan or of proper demand side management (DSM) strategies [10,13]. Moreover, the energy mangers can benefit from load profiles characterisation for exploiting anomaly detection initiatives [3].
- Energy Service Companies (ESCO) acting as building managers could benefit from the knowledge of typical building load profiles to develop energy savings and conservation measures.
- Transmission System Operators (TSOs) and Distribution System Operators (DSOs) may both benefit from energy profiles identification for the management of theirs grids and markets. In the case of smart electricity grids or

district heating installations a comprehensive analysis of energy demand curves can support demand management strategies aimed at improving the system balance [14].
- Policy makers may take advantage of load profile characterisation to identify which actions could have the highest impact for a specific group of consumers or buildings [11].

The identification of typical patterns in buildings is a matter of concern in different fields involving electrical and thermal energy demand in buildings. However, considering that the application of data analytics technologies to discover typical load profiles in buildings is relatively young, the implementation of robust frameworks is particularly desirable [15]. In this paper, a critical methodological framework based on the recent scientific literature to drive the extraction of typical load profiles from energy time series data is proposed. The process is discussed in relation to most important applications related to the classification of customers in smart city environment and single building/sub-system energy diagnosis. Moreover, an example of energy profile characterisation at individual building level for one-year energy consumption data related to a mechanical cooling/heating room is provided.

2. A robust recognition of load patterns in buildings

In Fig. 1 is shown a general framework for the extraction of typical load profiles and their use for different applications. To this purpose, data pre-processing is a preliminary phase aimed at organising time series data in an appropriate format to mine load energy patterns [3]. Moreover, a data preparation including data cleaning and data grouping is carried out in this phase. For example, buildings can be a-priori segmented into preliminary classes according to different properties (e.g., voltage, building type, weather conditions, building physical features etc.) to obtain more homogeneous groups that are simpler to be treated in a single framework. The second phase is performed at single building/customer level and it is aimed at extracting typical daily energy patterns according to profiles similarity generated under the same load conditions [11,13]. At this stage depending on whether a single or a group of buildings/customers are analysed, two different and alternative aims can be pursued.

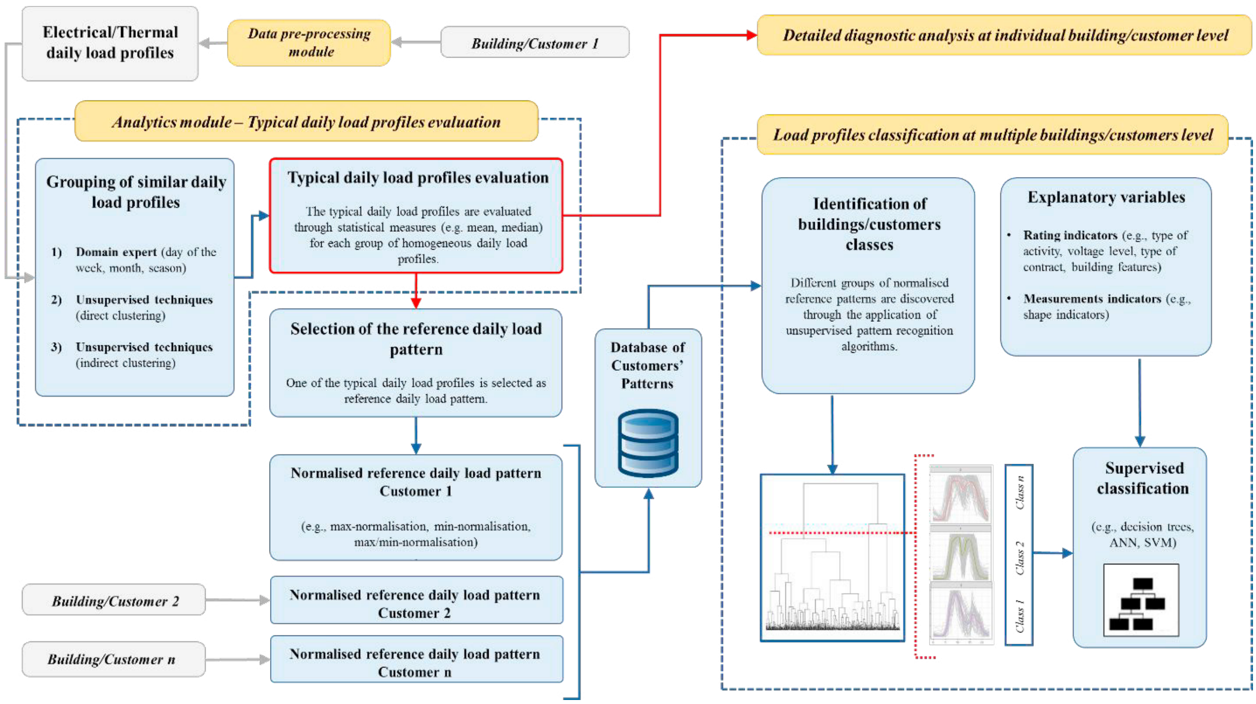

Fig. 1. General framework for load profiles characterisation at single building level and classification at multiple buildings/customers level.

When a group of buildings/customers are analysed, after the extraction of a reference load pattern for each of them a customers' classification is usually performed [16]. On the other hand, the discovery of typical load profiles can be also interpreted as a preliminary phase for supporting a detailed diagnostic analysis (e.g., energy demand prediction, fault detection and diagnosis (FDD), energy benchmarking) performed at sub-system or whole building level.

2.1. Data pre-processing

Data pre-processing is a fundamental task to prepare time series data for the load profiles recognition analysis. In a first step, the collected raw data in form of time series are analysed through different statistical methods to identify potential missing values and punctual outliers that must be replaced or removed [3]. In a second step, the original time series is chunked in fixed length windows in order to obtain constant time-scale based sub-sequences. A time-series is characterised by many individual data values, but in this case is viewed as a single object to process, representing the daily energy profile. This method, in the framework of time series analysis, could have potential drawbacks since patterns in some cases could be better characterised by different length of windows in the segmentation procedure. However, the daily scale is the most common time scale adopted in this field of investigation. The sub-sequences, representing the daily load profiles, are organized into a MxN matrix where M is the number of daily load profiles while N depends from the granularity of the collected measurements. Preliminary pattern recognition techniques (e.g., hierarchical clustering methods with average or single linkage [17]) can be used to detect daily load profiles presenting anomalous patterns that can negatively influence the performance of the analyses.

2.2. Typical daily load profiles evaluation

This phase of the framework is performed at individual building/customer level and it is aimed at identifying the typical daily load profiles that are the most representative of the energy behaviour. The number of profiles potentially extractable strongly depends by load conditions characterising the building/customer under analysis. To this purpose, data segmentation may be performed following (i) a domain expert based approach (ii) a data mining approach by using unsupervised techniques (iii) an indirect clustering through data reduction methods.

The typical profiles can be then evaluated through statistical measures (e.g. mean, median) calculated in each group of homogenous daily load profiles identified in the data segmentation phase. Generally, when the extraction of typical daily load profiles is performed at single building/customer level, data scaling is not carried out in order to preserve the effect of shapes as well as magnitudes in finding groups of similar objects.

Expert segmentation is a procedure completely driven by the domain knowledge of the analyst and it is aimed at generating subsets of daily load profiles that are supposed to be subjected to homogenous boundary conditions [18]. The typical load profiles are then usually calculated as the statistical mean of the objects in each subset. A common procedure consists in segmenting weekdays and weekend/holidays [19] due to the different activities carried out by a specific customer in these periods. Depending if the characterisation is or not performed on thermal sensitive energy patterns, the assumptions that need to be taken into account for an expert segmentation could be significantly different. In the case of characterization of thermal sensitive patterns, the climatic conditions (external and internal), occupancy pattern, building thermo-physical features and heating/cooling systems operation modes generate the existence of various load patterns (in terms of shape and magnitude), not always easily inferable. It is proven that also socio economics factors [20] and building final type [21] play a key role in the variation of load profiles. Especially in residential buildings, electrical (not thermal sensitive) and thermal load profiles do not have similar patterns [11,13]. For example space heating energy demand in households presents a typical two-peak daily pattern [20], instead of afternoon/evening peak pattern [1] that is typical of electricity demand. Moreover, daily load profiles for space heating exhibit small differences between weekdays and weekends making more reasonable a climate-driven segmentation considering a seasonal effect [13,21] (i.e., segmenting the warm periods from the cold periods).

In this perspective, in order to avoid the identification of noisy or not representative reference profiles more and more analysts rely on the application of unsupervised pattern recognition techniques such as cluster analysis [22]. Unlike the expert segmentation, cluster analysis allows load patterns to be identified in a not pre-determined time domain. In this way, robust and consistent reference profiles can be discovered.

A further approach for the data segmentation and profiles extraction relies on indirect clustering [23], where the object of clustering are features extracted from the load profiles. To this purpose, Symbolic Aggregate approXimation (SAX), can be used. SAX is based on the reduction of a time series through a piecewise process and then on its transformation into symbolic string [24]. The symbolic string represents a symbols subsequence preserving the important information of the original time series [25]. This approach offers the opportunity to discover *motif* (frequent symbols subsequence) and *discord* (infrequent symbols subsequence) for example fixing a threshold of the word frequency count for each identified pattern [26]. In this way the *discords* can be filtered out (e.g., for future detailed diagnostic investigation) while from the remaining set of *motifs* the typical daily load profiles are evaluated through cluster analysis on dataset of reduced dimensions. This approach makes it possible to find more robust typical profiles considering that sub-sequences based on both magnitude and shape similarity are discovered.

2.2.1 Load profiles characterization at multiple customers/buildings level

When a group of buildings/customers is analysed a classification process is usually performed. To this purpose a reference daily load pattern needs to be selected for each building/customer. In fact, the classification process involves a large number of customers making it a labour intensive and time consuming task. For this reason, in most of cases, it is necessary to extract only one representative load pattern from the set of typical daily load profiles of each building/customer evaluated through the previous phase. On the basis of the data segmentation, the representative load pattern usually corresponds to the typical profile in a specific time period or to the most populated cluster or to the most occurring motif. After the selection of the reference load pattern for each customer/building, data scaling is necessary in order to compare the different profiles between each other removing the effect of magnitude. Magnitude differences, resulting from different building design features (e.g. gross volume, floor area, installed power, etc.) or load conditions, can negatively affect the performance of pattern recognition algorithms in discovering similar shapes among daily load profiles. Scaling can be achieved through different approaches. Load profiles in the (0,1) range are obtained normalizing respect to a reference power e.g., the maximum value [27], mean value [12] or between minimum and maximum [22] values of the original daily load profiles. In other cases, a z-score normalisation can be also performed. Consequently, the representative normalised load patterns are processed in order to discover typical classes of customers/buildings and to classify them according to appropriate variables. The whole process consists of three different steps: (i) identification of *n* customer classes of buildings/customers, (ii) definition of the normalised reference load pattern for each customers' class (e.g. centroid) (iii) enrichment of the database with additional attributes (categorical or numerical) for each load profile to perform a supervised classification process.

The first step generally unfolds through the application of unsupervised pattern recognition techniques such as hierarchical or partitive cluster analysis. In the literature a plethora of combinations between aggregation techniques and distance measures were proposed. The best known methods include k-means algorithm [10,11,13,14,18,22,28], Fuzzy C-Means (FCM) [18,20,22,29], hierarchical clustering [18,22,30], Self-Organising Maps (SOM) [10,14,18] and Follow the Leader algorithm [17,27]. In particular, these techniques make it possible to handle a large amount of data in order to recognize the most significant load patterns. On the other hand, due to the necessity of setting in advance the clustering technique, the distance measure between objects and the number of clusters, this type of procedure is influenced by the subjective interpretation of the analyst. For this reason, several clustering adequacy indices (based on the measure of inter-cluster similarity and intra-cluster dissimilarity) have been proposed in the literature in order to drive the fine-tuning of the algorithms. In [18] were employed a series of adequacy measures such as Mean Index Adequacy (MIA), Clustering Dispersion Indicator (CDI), Scatter Index (SI), Variance Ratio Criterion (VRC) and Davies-Bouldin Index (DBI). In [10,22] the ratio of Within Cluster Sum of Squares to Between Cluster Variation (WCBCR) was indicated as the most suitable algorithm to evaluate the correct number of partitions of a clustering algorithm. Moreover the reliability of clustering techniques strongly depends by the selection of the distance measures particularly when time series sequences are compared [31]. For this reason, much effort was done in the recent years in finding robust distance measures (e.g. Dynamic Time Warping (DTW) [32]), to discover in a more robust way shapes similarity for a group of daily scaled energy patterns [31]. The outcome of this step is the definition of *n* classes of buildings/customers where the representative load patterns can be evaluated by calculating the centroid or medoid of the profiles grouped together. Subsequently, the customer class label is defined as a categorical dependent variable which can be predicted with a classification model, using additional attributes for the supervised process [19,33]. As shown in the framework (Fig. 1) the additional attributes to be taken into account in the classification, can be related to customers' rating information or based on field measurement campaign of energy patterns [34]. The rating indicators are related to the customers' contracts and to type of activity; they are generally used by the distribution companies to obtain a preliminary classification of their customers. These indicators are not shape sensitive and if used alone as predictors could be not able to provide a good characterization [34]. For this reason, indicators based on field measurements campaign can be employed in order to improve the accuracy of the classification model. For each customer these indicators are derived from measured daily representative load pattern providing compact information on the variation of the consumption over time. The indicators ranging in (0,1) are capable to highlight differences between customers behavior in different daily periods (e.g. night, lunch time) [34] considering that peak hours and off-peak hours could be different among customers. Once the predictors are selected, the methodological process goes through the development of a classification model. Also in this case the literature offers a wide set of suitable algorithms such as decision trees (e.g. C4.5, C5.0) [19,33]. Especially this kind of classifiers are used in customers classification processes due to their capability in handling both categorical and numerical variables and the high intelligibility of their output in terms of decision rules.

The described methodology represents a robust and useful tool to easily estimate for a new statistical object the representative load profile on the basis of the predicted membership to a specific class of customers/buildings. This opportunity in a liberalized energy market is highly desirable for suppliers, local and national authorities for implementing more effective energy management strategies also through targeted financial demand response programs [34–36] (e.g., Time Of Use tariff, Critical Peak Pricing, Real-Time Pricing). These programs are designed to be attractive for the consumers and at the same time profitable for the retailers. Indeed, the knowledge of customers' macro-behavior in energy consumption allows the distribution companies to better manage the grid operation and the interactions between energy consumption and production (e.g., implementation of energy storage solutions in district heating DH [37]). In this context, also the modification of customers energy use over time represents an effective demand side management strategy. The modification of a load profile (e.g., consuming less energy during peak hours or shifting the energy use to off-peak hours), allows the demand profile to be flat or in some cases to follow the generation pattern for grid stability purpose. For example, virtual thermal storage, through the modification of load profiles of a group of buildings served by a district heating network represents an effective way to increase the share of heat from cogeneration and renewable sources [38]. Moreover, the opportunity to classify specific groups of buildings/customers on the basis of their representative load patterns makes it possible to better address the current transition from large centralised generation plants to distributed ones that are capable to provide energy at a small-scale (e.g., neighborhood) when it is needed [39]. In fact, lack of knowledge about building energy use patterns currently represents the main barrier for fully exploiting the benefits of energy management also at micro grid level. In a dynamic smart city environment, the energy consumption patterns of the end-use customers need to be benchmarked also in order to assess the influence and the impact of DSM and DR initiatives over time [40].

2.2.2 Detailed diagnostic analysis at individual building/customer level

Following the framework reported in Fig. 1 the typical daily load profiles evaluation can also represents a preliminary analysis for supporting a detailed diagnostic procedure at individual building level. The knowledge of typical load profiles at single building/system level offers the opportunity to address complex emerging issues in building management such as building energy demand prediction, Fault Detection and Diagnosis (FDD), energy benchmarking, active demand response applications, exploitation of on-site renewable energy sources production. In fact, typical load profiles have a direct or indirect effect on the aforementioned applications. In [41] a motifs extraction based methodology was proposed to enhance the operation of a set of chillers serving a data center. Moreover, the sustainability impact was evaluated by means of useful metrics. In [31] a pattern recognition analysis based on a clustering algorithm (k-Shape) was performed in order to discover building energy consumption patterns. These patterns were further utilized to improve the accuracy and robustness of a forecasting model of energy consumption for ten institutional buildings in Singapore based on Support Vector Machines (SVM) algorithm. Also in [29] typical load profiles identification were used as a preliminary step in the development of a forecasting model for the electrical power demand of a supply fan of an Air Handling Unit (AHU). In detail, using a Fuzzy C-Means clustering algorithm, three subsets of homogeneous daily profiles (typical patterns) of the supply fan modulation were discovered while the atypical profiles were removed from the dataset. Then, for each subset a forecasting model combining Autoregressive Neural Network (ANN) and a physical model was built. The development of innovative robust multiple-diagnosis methodologies to automatically detect anomalous energy consumption [42] (profiles with shape/magnitude significantly different from the typical operation patterns) makes it possible to operate a continuous commissioning of the building, also defining rule-based strategies [43] to be implemented in the building energy management system. For building diagnosis purpose also the robust extraction of daily patterns of occupancy data or indoor environmental quality parameters can be extremely useful when it is correlated with energy usage patterns. For example, the occupancy profile can be associated to operation of air conditioning or lighting systems. On the other hand, the building energy usage patterns can be analysed in relation to different consuming components or sub-systems whose mutual interactions and correlations can be discovered by analysing their behaviour over time with a temporal mining approach. Temporal data mining can support the optimal operation of a building at multiple levels through the extraction of useful cross-sectional relationships between forcing variables and the actual energy consumption by performing a multivariate time series analysis. In some cases, can be beneficial to transform the typical daily load profiles to reduce datasets. In detail SAX representation can be employed to improve the execution of data mining algorithms increasing the computational efficiency [24]. In [26] a process based on SAX representation of time series to transform the original load profiles information into strings of characters was developed. This technique combined with motif discovery and clustering algorithms was employed to enhance diagnostic procedures of building energy usage. Moreover, the methodology proposed can be implemented in Fault Detection and Diagnosis (FDD) procedure and could provide relevant information for the calibration of simulation models. In [25] SAX and motif discovery were employed in combination with Temporal Association Rule Mining (TARM) to mine temporal correlations in Building Automation Systems (BAS) data. The extracted knowledge joined with domain expertise was helpful in identifying typical patterns and anomalies, estimating energy performance and detecting opportunities for energy conservation measures. As discussed above the robust recognition of energy profiles to perform energy diagnosis is particularly desirable. In the next section a robust methodological procedure conceived for this purpose is developed on a real case study.

3. Pattern recognition procedure for a single building/customer: a case study

The case study herewith presented deals with the application of a pattern recognition procedure applied to electrical consumption data releted to a heating/cooling mechanical room of Politecnico di Torino campus in Turin (Italy). The analysed system serves the main building of the campus including classrooms, central administration, bar and canteen for an overall floor area of more than 20,000 m^2. The system includes both hot and chilled water circuits of the building with the corresponding auxiliary pumping systems. The circulation pumps installed are different for the two circuits and have an overall designed electrical power of 120 kW. The hot water is produced through a district heating heat exchanger located in separate area of the campus. The chilled water, instead, is provided by two chillers (with a total design electrical power of 220 kW and a rated cooling capacity of 1120 kW) and a water to water reversible heat pump (with a design electrical power of 165 kW and a rated capacity of 590 kW in cooling mode). The two chillers and the heat pump are connected in parallel and the heat rejection is operated through a geothermal water source in a closed loop. The operation of chillers is controlled according to the cooling load of building to maintain supply/return temperature of the chilled water at 7/12 °C. The aggregated average electrical load data (chillers and auxiliaries) of the whole system for the year 2015 with a time interval of 15 minutes were analysed. The original time series data were chunked into daily sub-sequences that were organised into a 365 x 96 matrix. From the 365 daily load profiles, 70 were removed due to the presence of inconsistences and outliers. The remaining 286 daily load profiles were then analysed in order to characterise the operational patterns of the system under analysis. The open-source software R was used to perform the analysis.

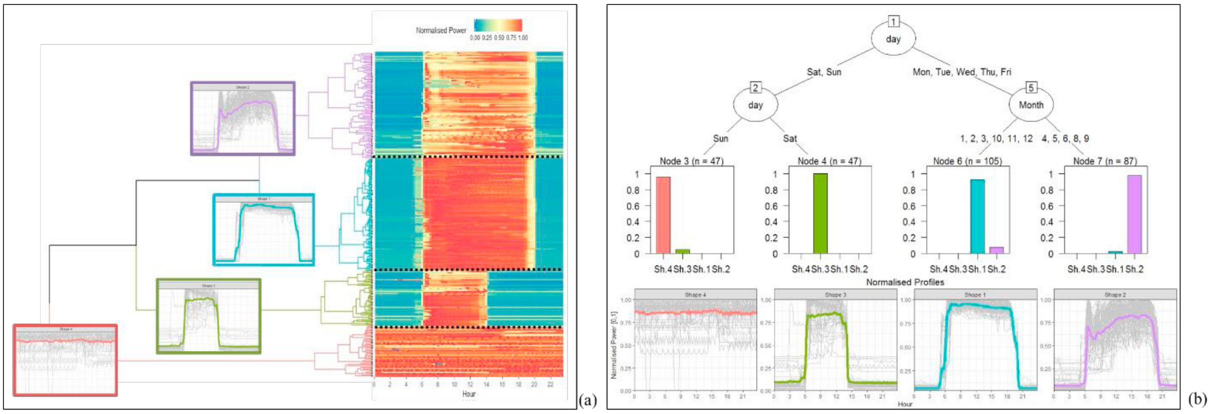

Fig. 2. (a) Shapes of load profiles extracted through hierarchical clustering; (b) Classification of profiles shapes.

In a first phase the daily load profiles were normalised in the (0,1) range on the maximum power value of each profile. A hierarchical clustering algorithm with Ward linkage method was then implemented to group the normalised load profiles. The Davies-Bouldin (DB) Index and the Silhouette Index were used as adequacy measures to identify the optimal number of partitions in a range between 2 and 10 clusters. In a first attempt 3 clusters were identified as the best partition. This first result produced 3 sets with different cardinality suggesting a further segmentation considering only the objects in the largest cluster. In this case Silhouette and DB indices suggested to split the daily load profiles in two further sub-clusters. In conclusion four different clusters labelled as “Shape 1”, “Shape 2”, “Shape 3” and “Shape 4” were obtained as shown in Fig. 2(a). The cluster labels were then used as dependent variable in a Classification and Regression Tree (CART) algorithm while *month* and *day of the week* were selected as explanatory variables. The result of the classification process produced a decision tree (Fig. 2(b)) capable to assign each shape to a specific period of the year (accuracy = 95.8%, precision = 96.4%, recall = 96.4%).This process provided a preliminary insight of the total power consumption patterns of the system under analysis. As shown in Fig. 2 (b), clusters “Shape 4” and “Shape 3” were found to be typical of Sundays and Saturdays, while clusters “Shape 1” and “Shape 2” of the working days of the winter and summer months respectively. Load profiles with Shape 1 are related to the period of the year where the electrical energy consumption of the system is due to the operation of the auxiliary pumping system of the hot water circuit. For this reason, load profiles in this cluster are characterised by a constant load between 6:00 and 20:00 with very small variations around the cluster centroid. On the other hand, cluster “Shape 2” is related to a period where chillers, auxiliary pumping system and geothermal water pumps were working under different conditions.

The normalized load profiles grouped in cluster “Shape 2” were rescaled to their original values to perform a further analysis. A segmentation of the energy profiles belonged in this cluster was performed and three different groups of daily profiles with magnitudes significantly different were discovered. In Fig. 3(a) are shown the centroids of these groups of profiles with the evidence of the relative standard deviation.

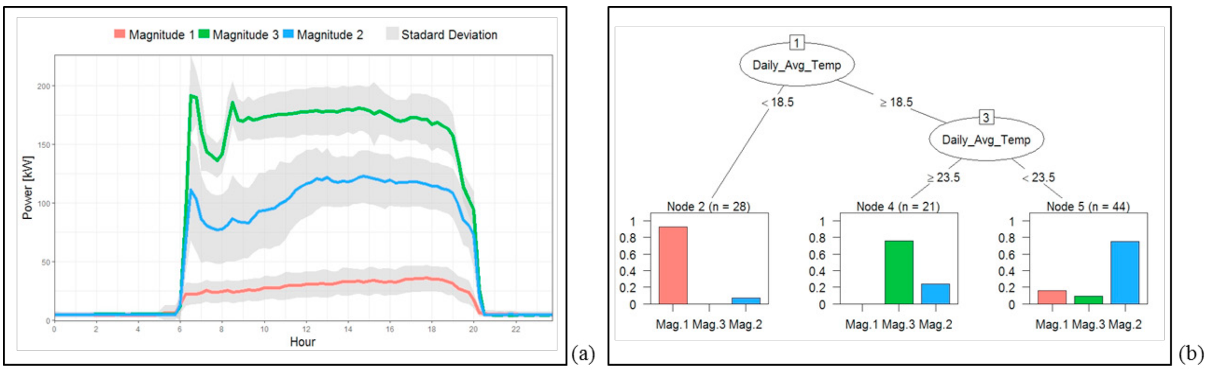

Fig. 3. (a) Load profile magnitudes extracted through hierarchical clustering; (b) Classification of profiles magnitudes.

The centroids of clusters labelled as “Magnitude 1”, “Magnitude 2” and “Magnitude 3” exhibit a mean value of the electrical power load during the time period 06:00–20:00 of 30 kW, 103 kW and 165 kW respectively. A classification process was then performed using as predictive variable the external air temperature. As shown in Fig. 3 (b), the profiles labelled as “Magnitude 1” occurred during days with a daily average temperature lower than 18.5 °C. The energy patterns of this cluster can be considered typical of mid-season when the operation of chillers were switched off but part of the auxiliaries (different from pumping system of the hot water circuit) were working. The clusters “Magnitude 2” and “Magnitude 3” grouped days of the warm season where different systems (chillers, heat pump and auxiliaries) were operating with external daily average temperature higher than 18.5 °C. Furthermore, the algorithm produced an additional split by classifying the load profiles of the warmest days (external daily average temperature ≥ 23.5°C) as “Magnitude 3”.

The approach presented in this case study was aimed at mining knowledge from energy consumption patterns and can be useful for detection potential deficiencies in control strategies or uncorrected operation of equipment and control system.

4. Conclusions

The great availability of data and platforms for managing them, contributed in the last decade to increase the interest in data-driven approach for facing emerging challenges in buildings in smart cities context. The development and implementation of generalised frameworks in this field of investigation represents a great issue to face, especially for combining different data mining techniques and couple them with building physics expertise for extracting useful information. In this perspective, in this work is proposed a general framework for implementing flexible and robust methodologies capable to drive the load profiles characterisation tasks at different levels (from sub-system to whole building) and at different scales (from single building to stock of buildings). In fact, lack of knowledge about building energy use patterns and how to promote their variation are currently the main barriers for fully utilizing the benefits of energy management in smart cities. The knowledge of energy consumption patterns at single system/building makes it possible to promote their optimisation through changes in energy demand, load shifting, the detection and diagnosis of anomalies related to uncorrected system operations or users’ behaviour. In this context future research will focus on exploring the potentialities of pattern recognition methodological process in handling inconsistent, infrequent, abnormal energy profiles considering a temporal diagnostic approach.

Acknowledgements

The authors express their gratitude to Living Lab of Politecnico di Torino for providing data for the case study presented in this work.

References

[1] Kmetty Z. Technical report - D4. 1 Load Profile Classification WP4 – Classification of EU residential energy consumers. 2017; 80.
[2] Capozzoli A, Cerquitelli T, Piscitelli MS. Chapter 11–Enhancing energy efficiency in buildings through innovative data analytics technologies, in: D. Ciprian, F. Xhafa (Eds.), Pervasive Comput, 2016; 353–389.
[3] Fan C, Xiao F, Yan C. A framework for knowledge discovery in massive building automation data and its application in building diagnostics. Autom Constr 2015; 50: 81–90.
[4] Molina-Solana M, Ros M, Ruiz MD, Gómez-Romero J, Martín-Bautista MJ. Data Science for Building Energy Management: a Review. Renew Sustain Energy Rev 2017; 70: 598–609.
[5] Yu Z, Fung BCM, Haghighat F. Extracting knowledge from building-related data - A data mining framework. Build Simul 2013; 6: 207–222.
[6] Capozzoli A, Piscitelli MS, Neri F, Grassi D, Serale G. A novel methodology for energy performance benchmarking of buildings by means of Linear Mixed Effect Model: The case of space and DHW heating of out-patient Healthcare Centres. Appl Energy 2016; 171: 592–607.
[7] Capozzoli A, Serale G, Piscitelli MS, Grassi D. Data mining for energy analysis of a large data set of flats. Proc Inst Civ Eng Eng Sustain 2017; 170: 3–18.
[8] Dudek G. Neural networks for pattern-based short-term load forecasting: A comparative study. Neurocomputing 2016; 205: 64–74.
[9] Fu TC. A review on time series data mining. Eng Appl Artif Intell 2011; 24: 164–181.
[10] Panapakidis IP, Papadopoulos TA, Christoforidis GC, Papagiannis GK. Pattern recognition algorithms for electricity load curve analysis of buildings. Energy Build 2014; 73: 137–145.

[11] Rhodes JD, Cole WJ, Upshaw CR, Edgar TF, Webber ME. Clustering analysis of residential electricity demand profiles. Appl Energy 2014; 135: 461–471.
[12] Spertino F, Chicco G, Ciocia A, Corgnati S, Leo P Di, Raimondo D. Electricity Consumption Assessment and PV System Integration in Grid-Connected Office Buildings. IEEE Int Conf Environ Electr Eng 2015; 255-260
[13] Do Carmo CMR, Christensen TH. Cluster analysis of residential heat load profiles and the role of technical and household characteristics. Energy Build 2016; 125: 171–180.
[14] Mcloughlin F, Duffy A, Conlon M. A clustering approach to domestic electricity load profile characterisation using smart metering data. Appl Energy 2015; 141: 190–199.
[15] Yu Z, Haghighat F, Fung BCM. Advances and challenges in building engineering and data mining applications for energy-efficient communities. Sustain Cities Soc 2016; 25: 33–38.
[16] Zhou K Le, Yang SL, Shen C. A review of electric load classification in smart grid environment. Renew Sustain Energy Rev 2013; 24: 103–110.
[17] Chicco G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. Energy 2012; 42: 68–80.
[18] Chicco G, Napoli R, Piglione F, Postolache P, Scutariu M, Toader C. Emergent electricity customer classification. IEE Proc Gener Transm Distrib 2005; 152: 164–172.
[19] Ramos S, Duarte JM, Duarte FJ, Vale Z. A data-mining-based methodology to support MV electricity customers' characterization. Energy Build 2015; 91: 16–25.
[20] Fernandes MP, Viegas JL, Vieira SM, Sousa JMC. Analysis of Residential Natural Gas Consumers using Fuzzy C-means Clustering. Proc 2016 IEEE Int Conf Fuzzy Syst 2016; 1484–1491.
[21] Gadd H, Werner S. Heat load patterns in district heating substations. Appl Energy 2013; 108: 176–183.
[22] Tsekouras GJ, Hatziargyriou ND, Dialynas EN. Two-stage pattern recognition of load curves for classification of electricity customers. IEEE Trans Power Syst 2007; 22: 1120–1128.
[23] Wang Y, Chen Q, Kang C, Zhang M, Wang K, Zhao Y. Load Profiling and Its Application to Demand Response: A Review. Tsinghua Sci Technol 2015; 20: 117–129.
[24] Lin J, Keogh E, Lonardi S, Chiu B. A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. Proc 8th ACM SIGMOD Work Res Issues Data Min Knowl Discov 2003; 2–11.
[25] Fan C, Xiao F, Madsen H, Wang D. Temporal knowledge discovery in big BAS data for building energy management. Energy Build 2015; 109: 75–89.
[26] Miller C, Nagy Z, Schlueter A. Automated daily pattern filtering of measured building performance data. Autom Constr 2015; 49: 1–17.
[27] Chicco G, Napoli R, Piglione F. Comparisons among clustering techniques for electricity customer classification. IEEE Trans Power Syst 2006; 21: 933–940.
[28] Ramos S, Duarte JMM, Soares J, Vale Z, Duarte FJ. Typical load profiles in the smart grid context a clustering methods comparison. IEEE Power Energy Soc Gen Meet 2012; 1-8
[29] Le Cam M, Daoud A, Zmeureanu R. Forecasting electric demand of supply fan using data mining techniques. Energy 2016; 101: 541–557.
[30] Jota PRS, Silva VRB, Jota FG. Building load management using cluster and statistical analyses. Int J Electr Power Energy Syst 2011; 33: 1498–1505.
[31] Yang J, Ning C, Deb C, Zhang F, Cheong D, Lee SE, Sekhar C, Tham KW. K-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. Energy Build 2017; 146: 27-37
[32] Petitjean F, Forestier G, Webb GI, Nicholson AE, Chen Y, Keogh E. Dynamic Time Warping Averaging of Time Series Allows Faster and More Accurate Classification. Proc - IEEE Int Conf Data Mining, ICDM 2015; 470–479.
[33] Figueiredo V, Rodrigues F, Vale Z, Gouveia JB. An electric energy consumer characterization framework based on data mining techniques. IEEE Trans Power Syst 2005; 20: 596–602.
[34] Chicco G, Napoli R, Postolache P, Scutariu M, Toader C. Customer characterization options for improving the tariff offer. IEEE Trans Power Syst 2003; 18: 381–387
[35] Siano P. Demand response and smart grids - A survey. Renew Sustain Energy Rev 2014; 30: 461–478.
[36] Wang K, Zhang M, Wang Z, Li R, Li F, Wu H. Time of use tariff design for domestic customers from flat rate by model-based clustering. Energy Procedia 2014; 61: 652–655.
[37] Kazas G, Fabrizio E, Perino M. Energy demand profile generation with detailed time resolution at an urban district scale: A reference building approach and case study. Appl Energy 2017; 193: 243–262.
[38] Verda V, Guelpa E, Sciacovelli A, Acquaviva A, Patti E. Thermal peak load shaving through users request variations in district heating systems. Int J Thermodyn 2016; 19: 168–176.
[39] Gelazanskas L, Gamage KAA. Demand side management in smart grid: A review and proposals for future direction. Sustain Cities Soc 2014; 11: 22–30.
[40] Benítez I, Quijano A, Díez J, Delgado I. Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. Int J Electr Power Energy Syst 2014; 55: 437–448.
[41] Patnaik D, Marwah M, Sharma R, Ramakrishnan N. Sustainable Operation and Management of Data Center Chillers Using Temporal Data Mining. Proc 15th ACM SIGKDD Int Conf Knowl Discov Data Mining 2009; 1305–1314.
[42] Capozzoli A, Lauro F, Khan I. Fault detection analysis using data mining techniques for a cluster of smart office buildings. Expert Syst Appl 2015; 42: 4324–4338.
[43] Peña M, Biscarri F, Guerrero JI, Monedero I, León C. Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach. Expert Syst Appl 2016; 56: 242–255.